Sequential Adaptive Learning for Speaker Verification

Jun Wang

CSDLT, RIIT, THU

2013-03-01
Outline

1. Introduction
2. Sequential Adaptive Learning
3. Experiments
4. Conclusion
5. Reference
Introduction

- GMM-UBM-based speaker verification heavily relies on a well trained UBM.
- In practice, it is often difficult to collect sufficient channel-matched data to train a fully consistent UBM.
- A multitude of research has been proposed to address channel mismatch or session variation.
Introduction

Within the GMM-UBM framework,
- feature transform [3, 4, 5];
- model compensation [6, 7];
- score normalization [1, 8];
- factor analysis [9,10] and its simple algorithm implementation [11];
- various feature and model compensation approaches [12];

Besides GMM-UBM,
- [13] proposed to reduce channel impact in neural network;
we propose a sequential adaptive learning approach to the channel mismatch problem.

By this approach, the UBM and speaker models are updated sequentially and gradually, finally converging to the new or dynamic channel with a large amount of enrollments.
Sequential Adaptive Learning

➢ **Review of MAP estimation**

The objective function:

\[ L(\mu, \sigma) = \log P(\mu, \sigma | X) \]

\[ \propto \sum_i \log \{ \mathcal{N}(x_i; \mu, \sigma) P(\mu, \sigma) \}. \]

Maximizing this objective leads to the following MAP estimation:

\[ \mu = \frac{\sum_i x_i + \frac{\sigma}{\hat{\sigma}} \hat{\mu}}{N + \frac{\sigma}{\hat{\sigma}}} \quad \text{(1)} \]
Sequential Adaptive Learning

- **Review of MAP estimation**

  When extending to GMM:

  \[
  r_i(c) = \frac{\mathcal{N}(x_i; \mu_c, \sigma_c)}{\sum_m \mathcal{N}(x_i; \mu_m, \sigma_m)}.
  \]

  \[
  (2)
  \]

  Define the following sufficient statistics:

  \[
  r_c = \sum_i r_i(c)
  \]

  \[
  z_c = \sum_i r_i(c) x_i,
  \]

  \[
  (3)
  \]

  \[
  (4)
  \]

  the MAP estimation is given by:

  \[
  \hat{\mu}_c = \frac{z_c + \frac{\sigma}{\hat{\sigma}} \hat{\mu}}{r_c + \frac{\sigma}{\hat{\sigma}}}
  \]

  \[
  (5)
  \]
Sequential Adaptive Learning

Sequential UBM adaptation

Motivation of sequential UBM adaptation:
Use the new enrollment speech data to update the UBM. We start from a ‘pool and re-estimation’ procedure.

\[
\mu_c = \frac{z_c + \hat{z}_c}{r_c + \hat{r}_c}
\]

\[
= \frac{z_c + \hat{r}_c \hat{\mu}_c}{r_c + \hat{r}_c}
\]

\[
\hat{r}_c = \frac{\sigma}{\hat{\sigma}}.
\]
Sequential Adaptive Learning

Fig. 1. Sequential UBM MAP adaptation.
Sequential Adaptive Learning

Sequential Speaker Model adaptation

Firstly, we need to save sufficient statistics for each speaker which are defined in equation (3) and (4).

When a new enrollment occurs, sequential UBM adaptation is used to train a new UBM, then we use the new UBM to update each speaker model according to it’s sufficient statistics.

\[
\mu = \frac{z_c + \frac{\sigma}{\hat{\sigma}} \hat{\mu}_n}{r_c + \frac{\sigma}{\hat{\sigma}}} \tag{9}
\]
Sequential Adaptive Learning

Fig2. Sequential adaptive learning
Experiments

- We conduct the experiments on a time-varying database [14].
- We start the experiments with two initial UBMs.
- The verification performance is evaluated in terms of equal error rates (EER).
Experiments

- Sequential UBM adaptation experiment.

<table>
<thead>
<tr>
<th></th>
<th>$k_s = 0.5$</th>
<th>$k_s = 1$</th>
<th>$k_s = 2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>UBM (baseline)</td>
<td>11.75</td>
<td>11.92</td>
<td>11.75</td>
</tr>
<tr>
<td>SUBM ($k_u = 90$)</td>
<td>12.69</td>
<td>12.42</td>
<td>12.05</td>
</tr>
<tr>
<td>SUBM ($k_u = 180$)</td>
<td>11.67</td>
<td>11.20</td>
<td>11.47</td>
</tr>
<tr>
<td>SUBM ($k_u = 270$)</td>
<td>11.19</td>
<td>11.05</td>
<td>11.07</td>
</tr>
<tr>
<td>SUBM ($k_u = 360$)</td>
<td><strong>11.12</strong></td>
<td><strong>11.02</strong></td>
<td><strong>11.05</strong></td>
</tr>
</tbody>
</table>

Table 1. Results with $\text{UBM}_\alpha$ as the initial.
Experiments

Sequential UBM adaptation experiment.

<table>
<thead>
<tr>
<th></th>
<th>$k_s=0.5$</th>
<th>$k_s=1$</th>
<th>$k_s=2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>UBM (baseline)</td>
<td>10.04</td>
<td>10.22</td>
<td>10.44</td>
</tr>
<tr>
<td>SUBM ($k_u=90$)</td>
<td>9.36</td>
<td>9.20</td>
<td>8.83</td>
</tr>
<tr>
<td>SUBM ($k_u=180$)</td>
<td>8.77</td>
<td>8.88</td>
<td>8.82</td>
</tr>
<tr>
<td>SUBM ($k_u=270$)</td>
<td>8.72</td>
<td>8.84</td>
<td>8.79</td>
</tr>
<tr>
<td>SUBM ($k_u=360$)</td>
<td>8.73</td>
<td>8.82</td>
<td>8.79</td>
</tr>
</tbody>
</table>

Table 2. Results with UBM$_b$ as the initial.
Experiments

- Sequential Adaptive Learning experiment.

<table>
<thead>
<tr>
<th></th>
<th>$k_s=0.5\varphi$</th>
<th>$k_s=1\varphi$</th>
<th>$k_s=2\varphi$</th>
</tr>
</thead>
<tbody>
<tr>
<td>UBM (baseline)</td>
<td>11.75%\varphi</td>
<td>11.92%\varphi</td>
<td>11.75%\varphi</td>
</tr>
<tr>
<td>SUBM ($k_u=90$)</td>
<td>9.34%\varphi</td>
<td>8.90%\varphi</td>
<td>8.65%\varphi</td>
</tr>
<tr>
<td>SUBM ($k_u=180$)</td>
<td>9.37%\varphi</td>
<td>9.04%\varphi</td>
<td>8.95%\varphi</td>
</tr>
<tr>
<td>SUBM ($k_u=270$)</td>
<td>9.47%\varphi</td>
<td>9.22%\varphi</td>
<td>9.08%\varphi</td>
</tr>
<tr>
<td>SUBM ($k_u=360$)</td>
<td>9.52%\varphi</td>
<td>9.35%\varphi</td>
<td>9.24%\varphi</td>
</tr>
<tr>
<td>SUBM ($k_u=540$)</td>
<td>9.74%\varphi</td>
<td>9.54%\varphi</td>
<td>9.54%\varphi</td>
</tr>
</tbody>
</table>

Table 3. Sequential adaptive learning with UBM\varphi
Experiments

- Sequential Adaptive Learning experiment.

<table>
<thead>
<tr>
<th></th>
<th>System EER&lt;sub&gt;k_d&lt;/sub&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>k_d=0.5&lt;sup&gt;φ&lt;/sup&gt;</td>
</tr>
<tr>
<td>UBM (baseline)</td>
<td>10.04%&lt;sup&gt;φ&lt;/sup&gt;</td>
</tr>
<tr>
<td>SUBM (k_u=90)</td>
<td>6.78%&lt;sup&gt;φ&lt;/sup&gt;</td>
</tr>
<tr>
<td>SUBM (k_u=180)</td>
<td>6.78%&lt;sup&gt;φ&lt;/sup&gt;</td>
</tr>
<tr>
<td>SUBM (k_u=270)</td>
<td>6.92%&lt;sup&gt;φ&lt;/sup&gt;</td>
</tr>
<tr>
<td>SUBM (k_u=360)</td>
<td>7.10%&lt;sup&gt;φ&lt;/sup&gt;</td>
</tr>
<tr>
<td>SUBM (k_u=540)</td>
<td>7.43%&lt;sup&gt;φ&lt;/sup&gt;</td>
</tr>
</tbody>
</table>

*Table 4. Sequential adaptive learning with UBM*
Experiments

➢ Quality of sequential UBM.

Fig3. Quality of sequentially adapted UBM.
Conclusion

- By adapting an initial UBM with a strong prior whenever a new enrollment is available, the UBM learns and converges to the working channel gradually, leading to improved verification performance.
- Use the new UBM to update each speaker model according to its sufficient statics, leading to improved verification performance.
- In our experiments, this sequential approach provides relative EER reduction of 24.1% and 34.9% for two mismatched UBMs, respectively.
Reference


Reference


Thanks
Q&A.